Towards multiscale high performance QM/MM simulations : a dialogue between Mathematics and Theoretical Chemistry

Louis Lagardère, Filippo Lipparini, Benjamin Stamm, Eric Cancès, Benedetta Mennucci, Yvon Maday, Jean-Philip Piquemal

Labex CALSIMLAB

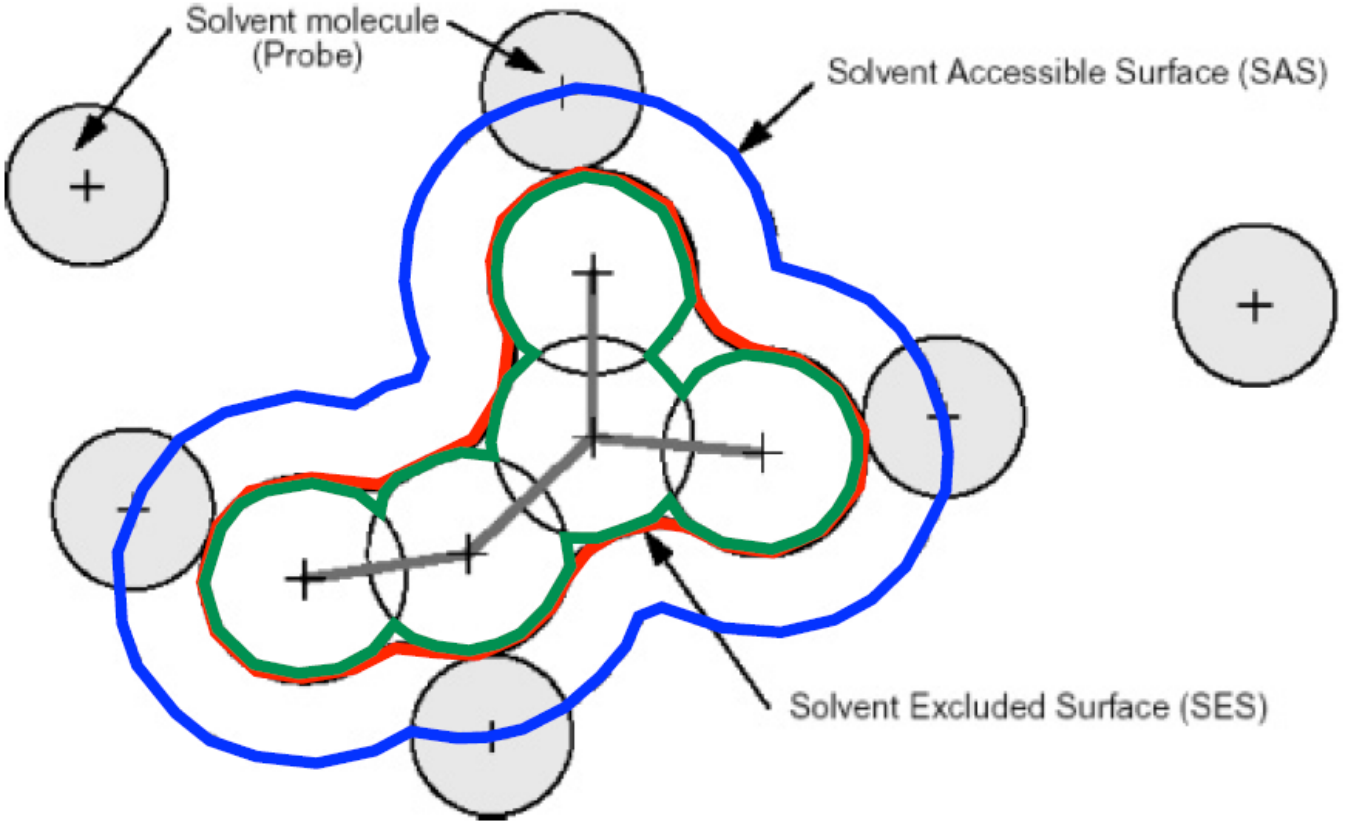
Outline

Motivations

1. Difficulties
   1. Huge number of different mathematical problems to tackle
   2. Very limited complete mathematical description
   3. Absence of “true” mathematical proof for lots of methods
   4. Sometime 20 years can last before a working technique is really understood
1. Objectives: a real synergy between Maths and Chemistry
   1. Developpement of a multiscale hybrid approach (QM/MMPol/Solvent)
   2. Applications to complex/large systems required
   3. Importance of massivelly parallel implementations (Equipex/mesocentres - centres nationaux GENCI)

COnductor Screening MOdel

- The solute is accommodated in a hollow cavity (Ω)



- An infinite structureless continuum occupies the rest of the space ($\mathbb{R}^3 \setminus \Omega$)
- The continuum is treated as a metal (conductor); the results are scaled with and empirical factor to account for its dielectric nature
- The electrostatic interaction between the solvent and the solute is treated *exactly*, by numerically solving the Poisson equation
- Other interactions can be treated with empirical expressions

The mathematical problem

Hypotheses:

- The cavity Ω is a compact, open subset of $\mathbb{R}^3$ with a regular enough boundary $\Gamma = \partial\Omega$
- The support of the solute's density of charge ρ is entirely contained in Ω

The solvation energy is

$$E^s = \frac{1}{2f(\varepsilon)} \int_\Omega d\mathbf{r}\rho(\mathbf{r})W(\mathbf{r})$$

- $f(\varepsilon)$ is an empirical scaling factor
- W is the *reaction-field potential*, solution to the COSMO equation.

Reaction potential:

It is the contribution to the total potential φ due to the presence of the conductor:

$$\varphi = \Phi + W$$

where Φ is the solute's potential in vacuo.

Total potential:

Solves Poisson's equation in Ω with metallic boundary conditions on Γ:

$$\begin{cases} \nabla^2\varphi = -4\pi\rho & \text{in } \Omega \\ \varphi = 0 & \text{on } \Gamma \end{cases} \quad \Rightarrow \quad \begin{cases} \nabla^2 W = 0 & \text{in } \Omega \\ W = -\Phi & \text{on } \Gamma \end{cases}$$

Integral representation:

$W \in H^1(\Omega)$ is a *single-layer potential*: it can be represented by an apparent surface charge (ASC) $\sigma \in H^{-1/2}(\Gamma)$ induced on Γ

$$\forall\, \mathbf{s} \in \Gamma\, W(\mathbf{s}) = \int_\Gamma \frac{\sigma(\mathbf{s}')}{|\mathbf{s}-\mathbf{s}'|}$$

We can recast the COSMO equation...

as an integral equation:

$$\forall\, \mathbf{s} \in \Gamma \int_\Gamma \frac{\sigma(\mathbf{s}')}{|\mathbf{s}-\mathbf{s}'|} := (\mathcal{S}\sigma)(\mathbf{s}) = -\Phi(\mathbf{s})$$

$$E^s = \frac{1}{2f(\varepsilon)} \int_\Omega d\mathbf{r}\rho(\mathbf{r})W(\mathbf{r}) = \frac{1}{2f(\varepsilon)} \int_\Gamma \sigma(\mathbf{s})\Phi(\mathbf{s})d\mathbf{s}$$

- Analytical solution exist only for very simple surfaces (sphere, ellipsoid...)
- For general, molecule-shaped surfaces, a numerical discretization is mandatory

Classical discretization for the PCM equations

Original problem (in its weak formulation):

$$\begin{cases} \text{seek } \sigma \in H^{-1/2}(\Gamma) \text{ such that} \\ \forall u \in H^{-1/2}(\Gamma) \quad \langle u, \mathcal{S}\sigma \rangle = -\langle u, \Phi \rangle \end{cases}$$

Galerkin approximation

The COSMO problem can be solved in a finite dimensional approximation space V_h

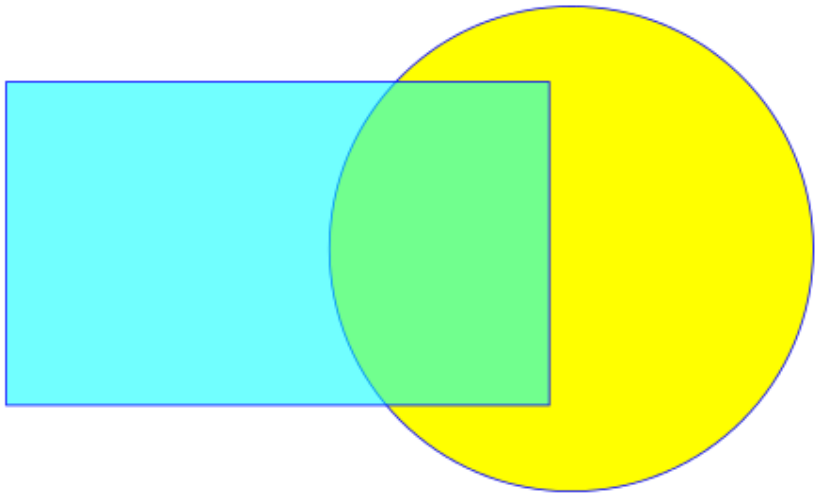
$$\begin{cases} \text{seek } \sigma_h \in V_h \text{ such that} \\ \forall u_h \in V_h \quad \langle u_h, \mathcal{S}_h\sigma_h \rangle = -\langle u_h, \Phi_h \rangle \end{cases}$$

By expanding σ_h in a basis of V_h...

one gets a linear system

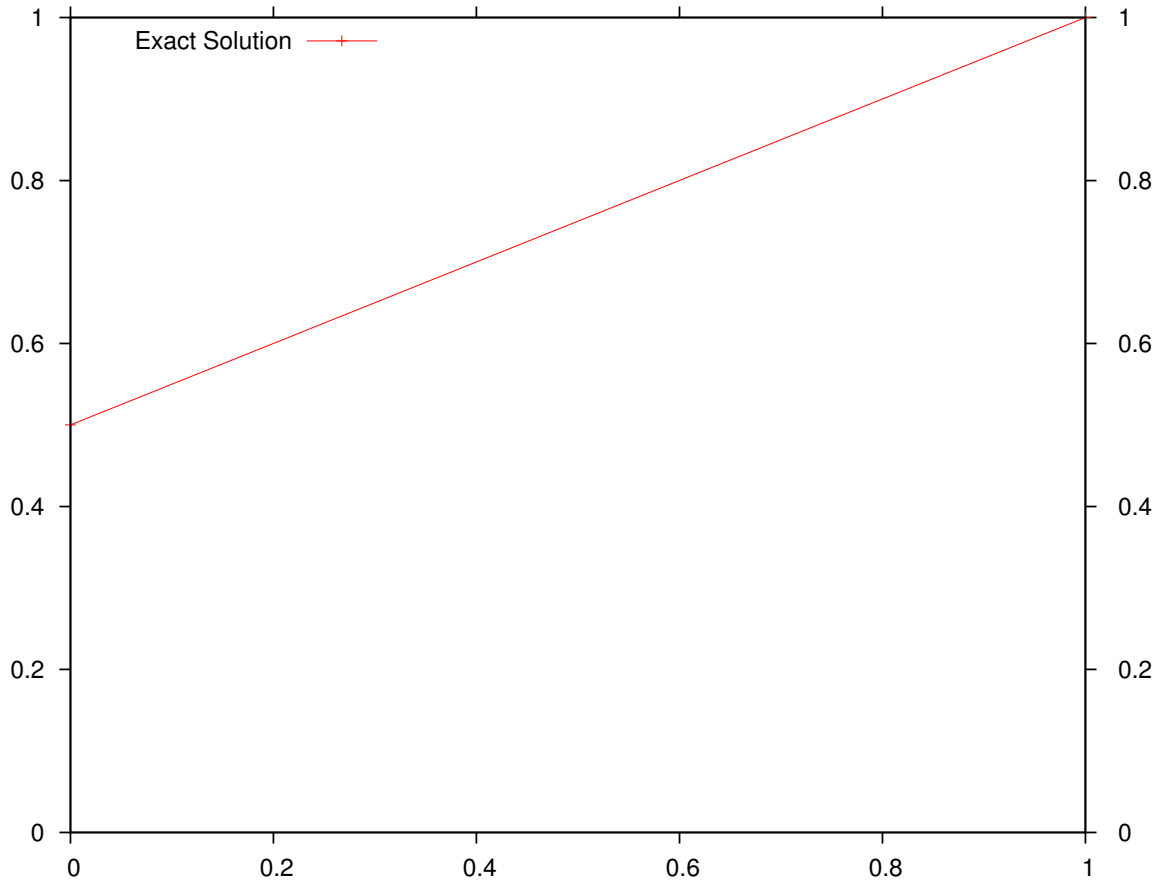
$$\mathbf{Sq} = -\mathbf{V}$$

Schwarz's Domain Decomposition Method



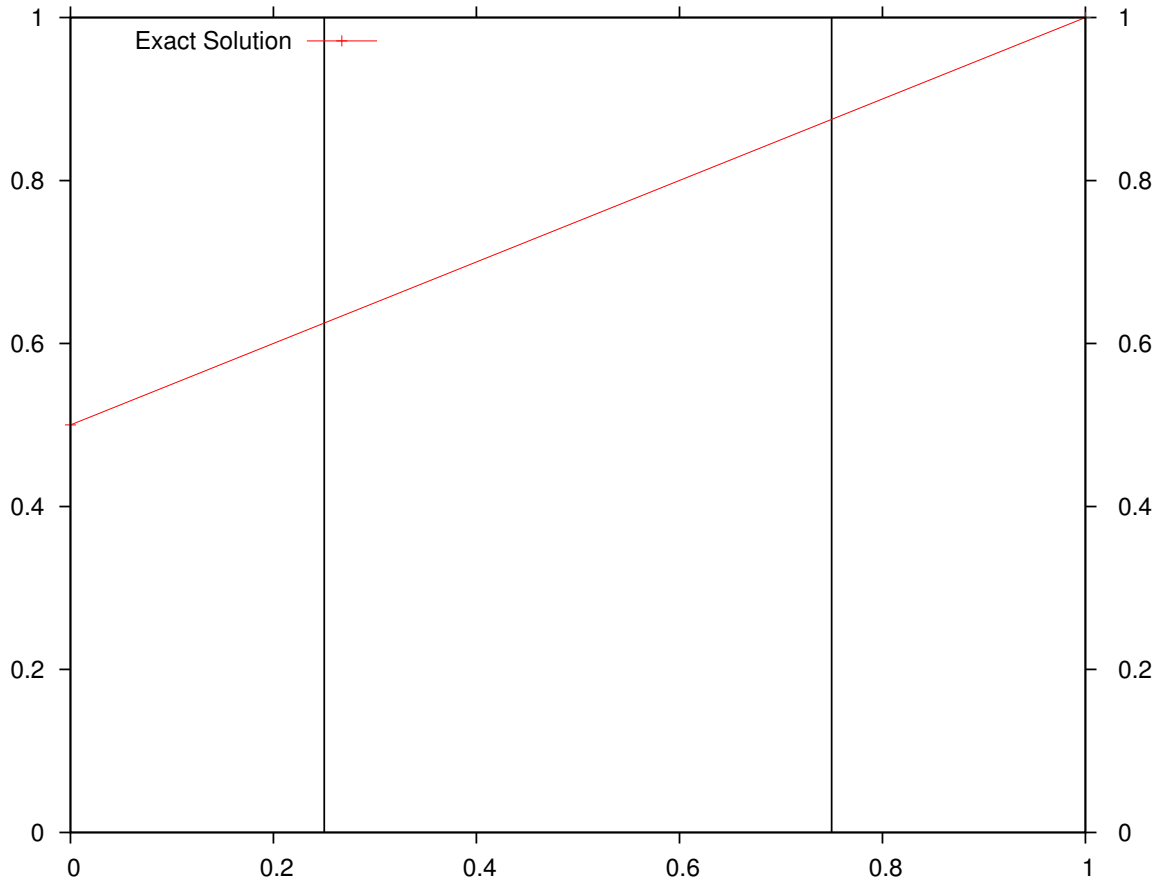
$$f'' = 0 \text{ in }]0,1[; \quad f(0) = 0.5, f(1) = 1.0$$

$$\Omega =]0,1[$$



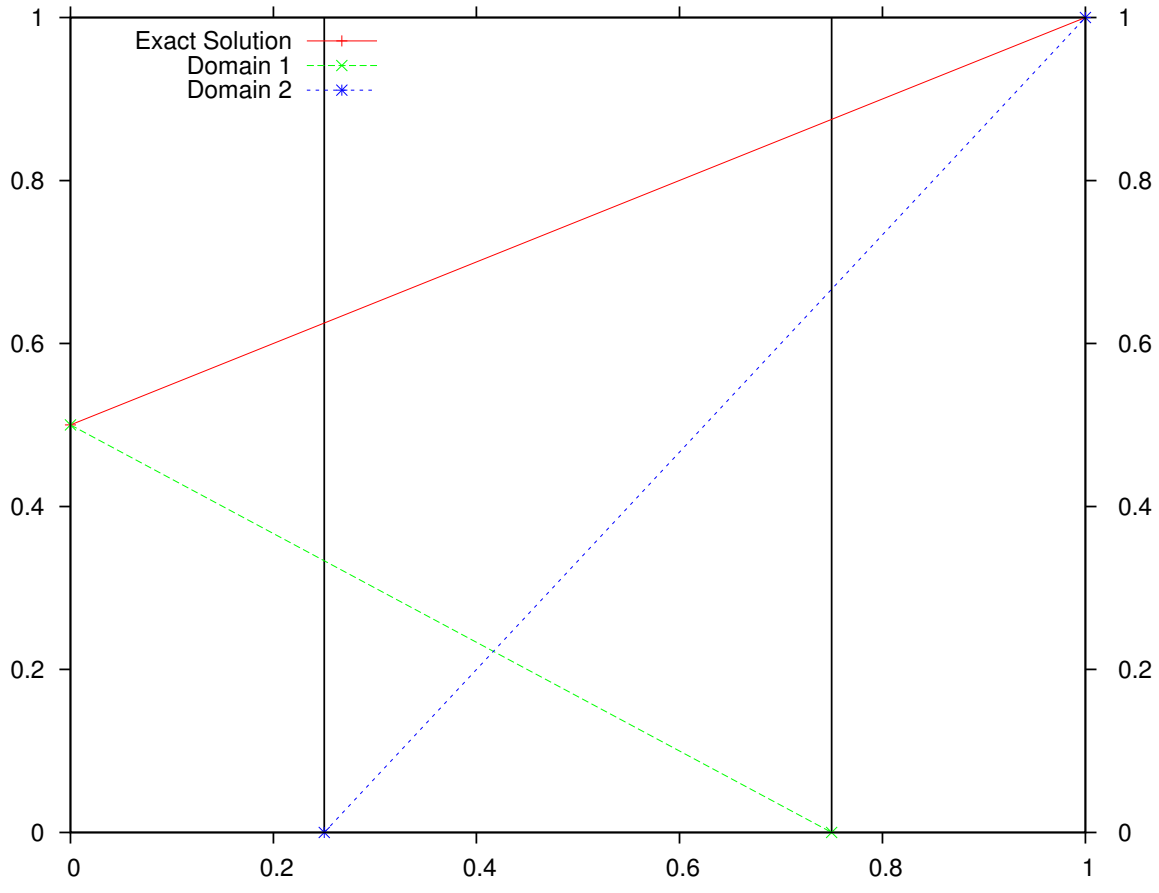
$$f'' = 0 \text{ in }]0,1[; \quad f(0) = 0.5, f(1) = 1.0$$

$$\Omega_1 =]0, 0.75], \quad \Omega_2 = [0.25, 1[$$



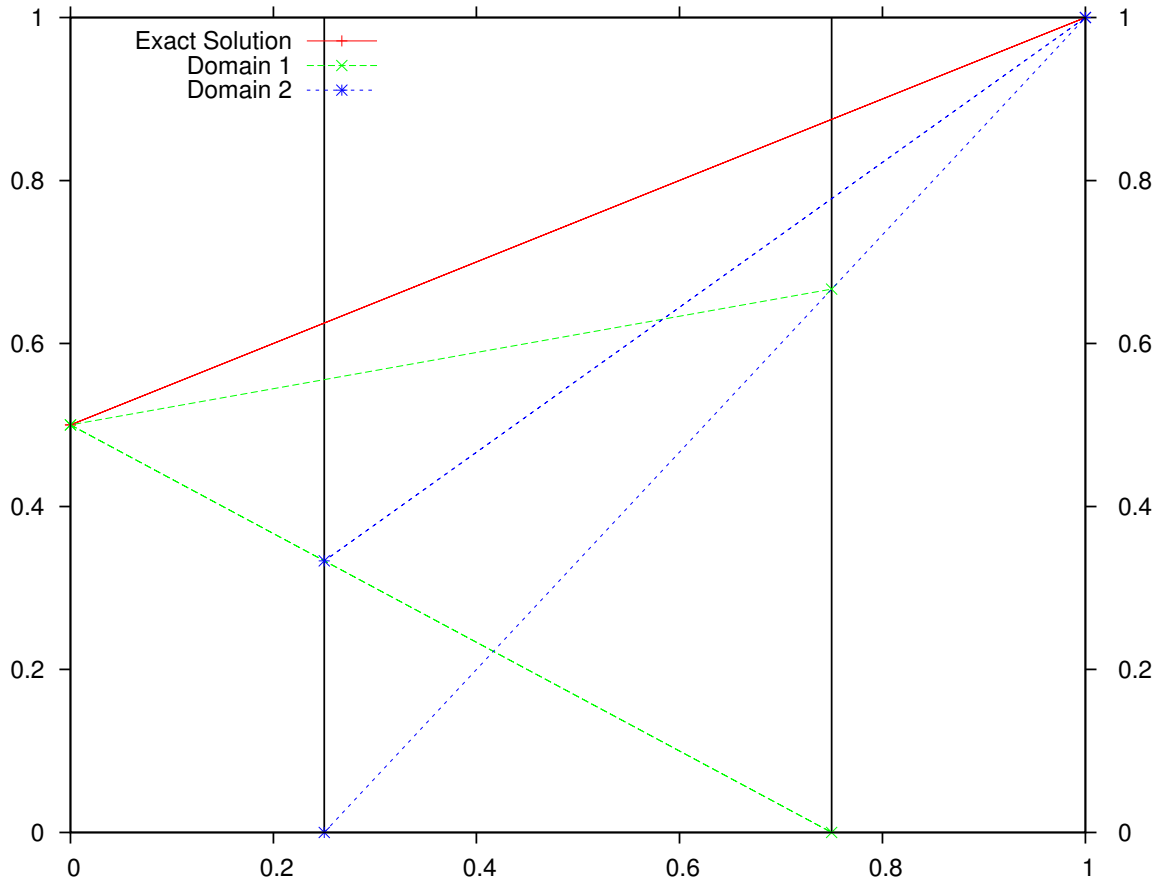
$$f'' = 0 \text{ in }]0,1[; \quad f(0) = 0.5, f(1) = 1.0$$

$$\Omega_1 =]0, 0.75], \quad \Omega_2 = [0.25, 1[$$



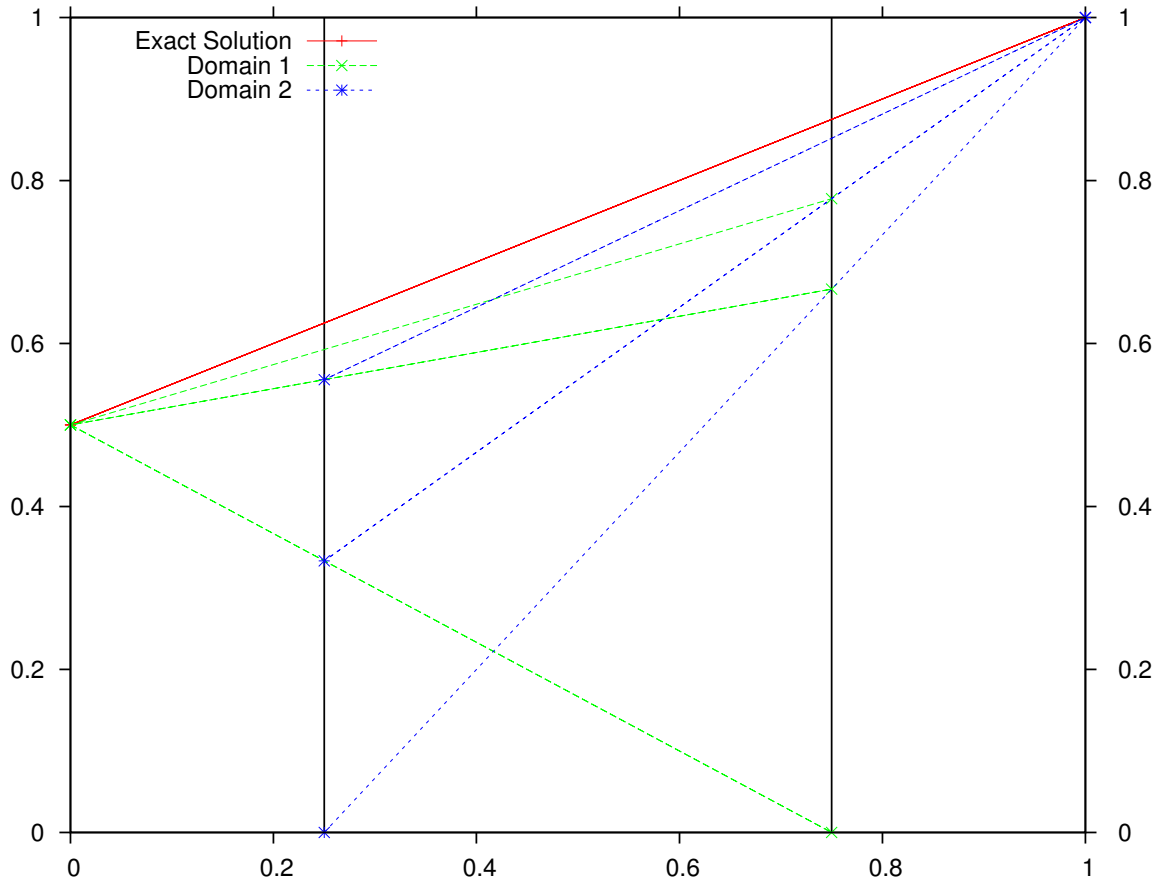
$$f'' = 0 \text{ in }]0,1[; \quad f(0) = 0.5, f(1) = 1.0$$

$$\Omega_1 =]0, 0.75], \quad \Omega_2 = [0.25, 1[$$



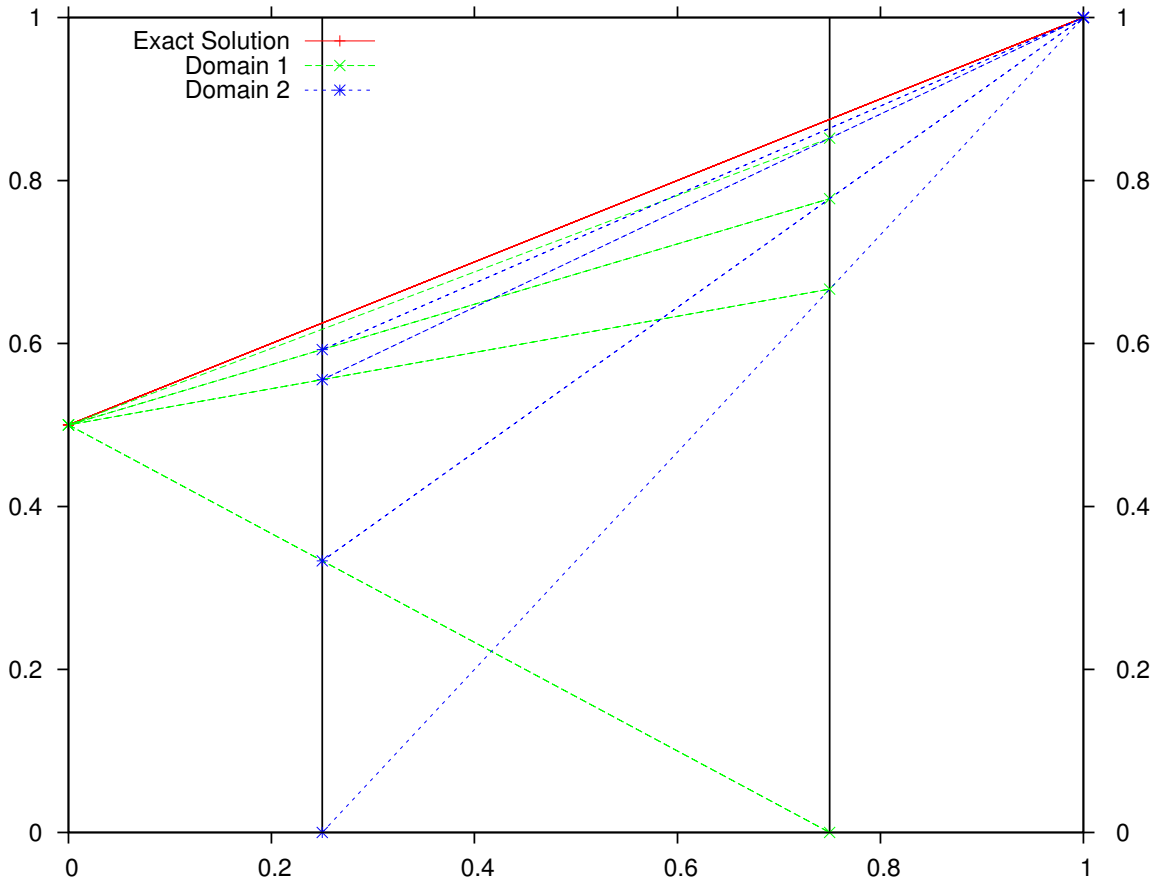
$$f'' = 0 \text{ in }]0,1[; \quad f(0) = 0.5, f(1) = 1.0$$

$$\Omega_1 =]0, 0.75], \quad \Omega_2 = [0.25, 1[$$



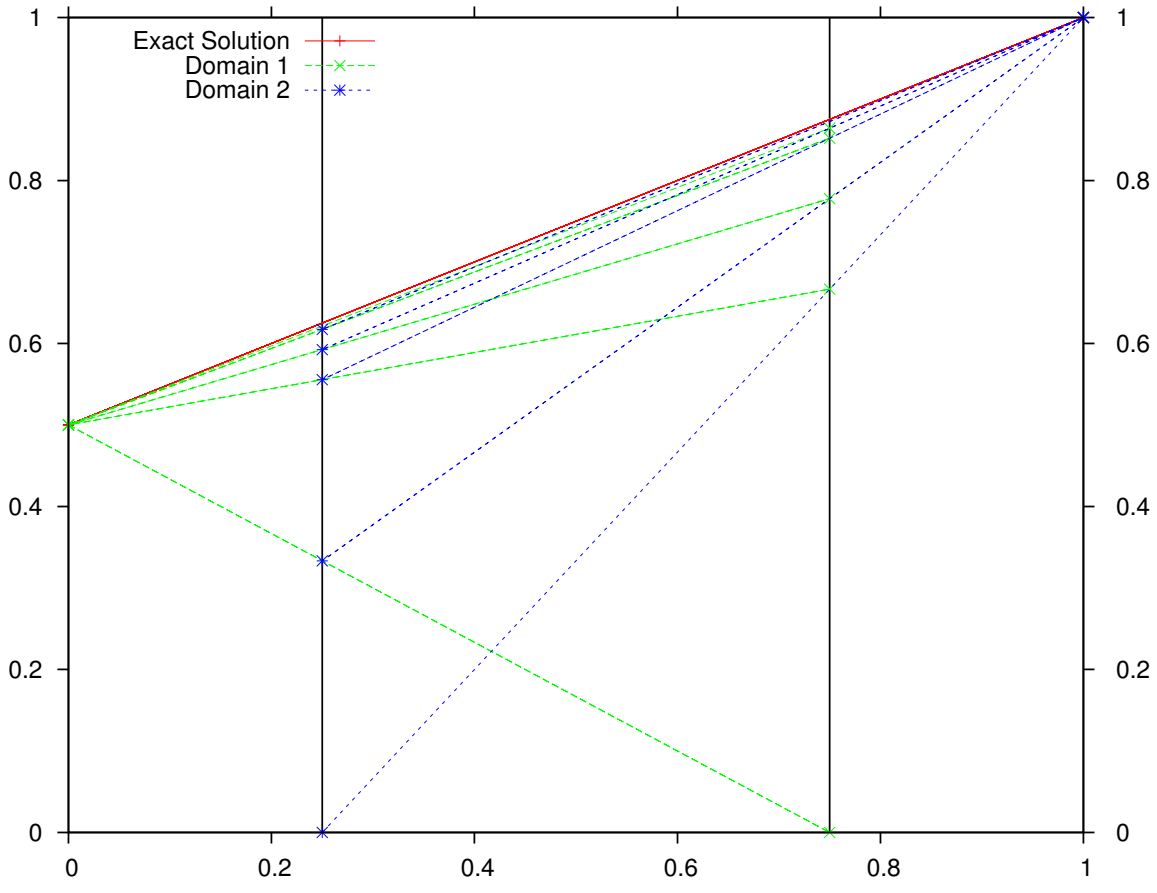
$$f'' = 0 \text{ in }]0,1[; \quad f(0) = 0.5, f(1) = 1.0$$

$$\Omega_1 =]0, 0.75], \quad \Omega_2 = [0.25, 1[$$

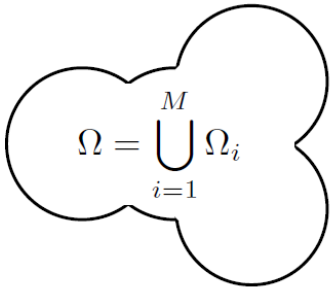


$$f'' = 0 \text{ in }]0,1[; \quad f(0) = 0.5, f(1) = 1.0$$

$$\Omega_1 =]0, 0.75], \quad \Omega_2 = [0.25, 1[$$

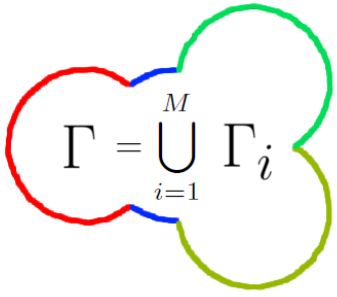


Schwarz's Domain Decomposition method



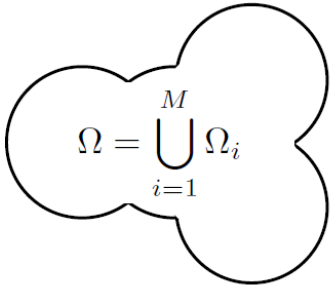
$$\mathcal{S}\sigma = -\Phi$$

$$\downarrow ?$$

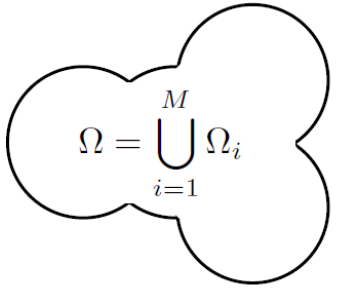


$$\mathcal{S}_i\sigma_i = g_i$$

Schwarz's Domain Decomposition method



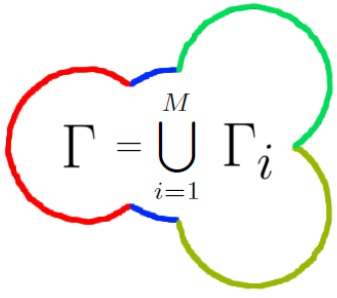
$\rightarrow$



$$\mathcal{S}\sigma = -\Phi$$

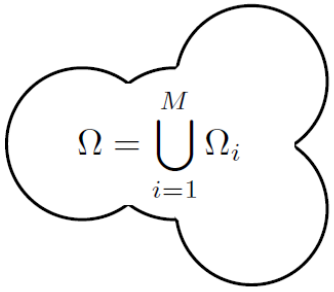
$$\downarrow ?$$

$$\begin{cases} \nabla^2 W = 0 & \text{in } \Omega \\ W = -\Phi & \text{on } \Gamma \end{cases}$$

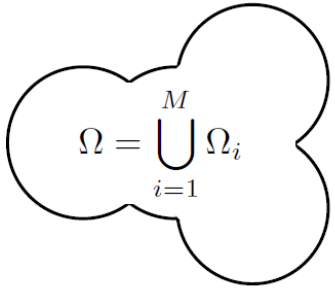


$$\mathcal{S}_i \sigma_i = g_i$$

Schwarz's Domain Decomposition method



$\rightarrow$

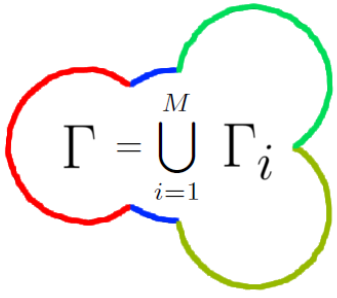


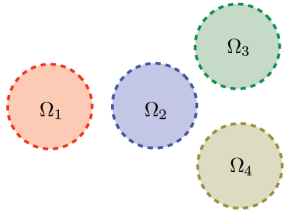
$$\mathcal{S}\sigma = -\Phi$$

$\downarrow ?$

$$\begin{cases} \nabla^2 W = 0 & \text{in } \Omega \\ W = -\Phi & \text{on } \Gamma \end{cases}$$

$\downarrow$

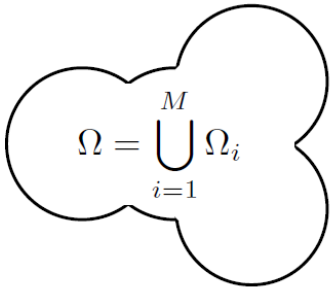




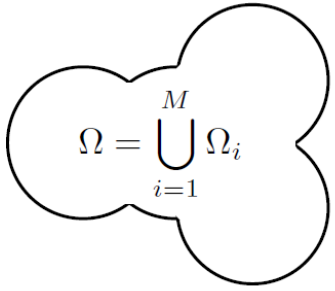
$$\mathcal{S}_i \sigma_i = g_i$$

$$\begin{cases} \nabla^2 W_i = 0 & \text{in } \Omega_i \\ W_i = g_i & \text{on } \Gamma_i = \partial\Omega_i \end{cases}$$

Schwarz's Domain Decomposition method



$\rightarrow$

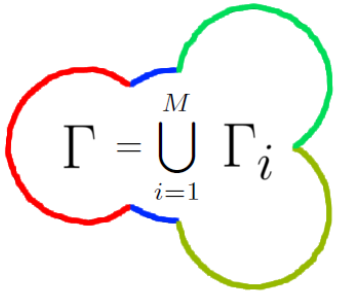


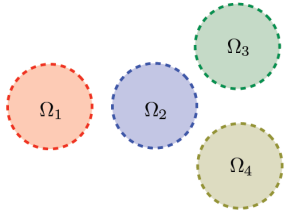
$$\mathcal{S}\sigma = -\Phi$$

$\downarrow ?$

$$\begin{cases} \nabla^2 W = 0 & \text{in } \Omega \\ W = -\Phi & \text{on } \Gamma \end{cases}$$

$\downarrow$





$\leftarrow$

$$\mathcal{S}_i \sigma_i = g_i$$

$$\begin{cases} \nabla^2 W_i = 0 & \text{in } \Omega_i \\ W_i = g_i & \text{on } \Gamma_i = \partial\Omega_i \end{cases}$$

Discretization

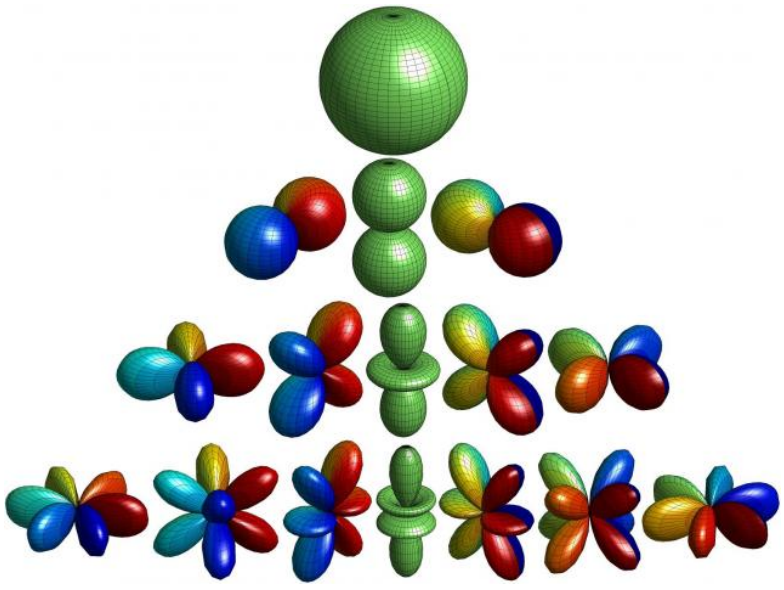


Figure : For each sphere, a set of spherical harmonics up to degree (angular momentum) N (usually 10 is a safe choice) is used to represent σ_i and the operators

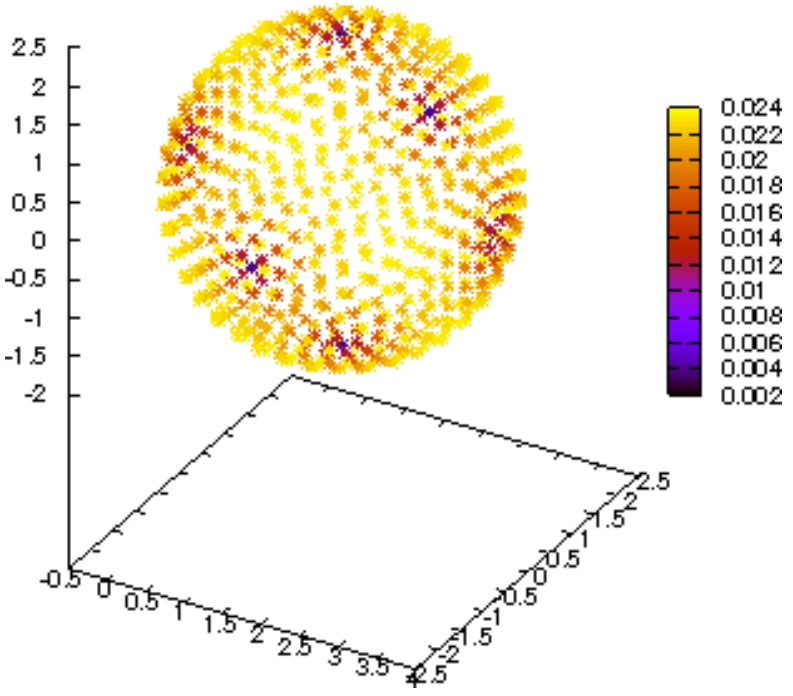


Figure : On each sphere, the numerical integrals are carried out with Lebedev quadrature, using a set $\{\mathbf{s}_n, w_n\}$ of N_g points and weights (~ 300 is usually a safe choice)

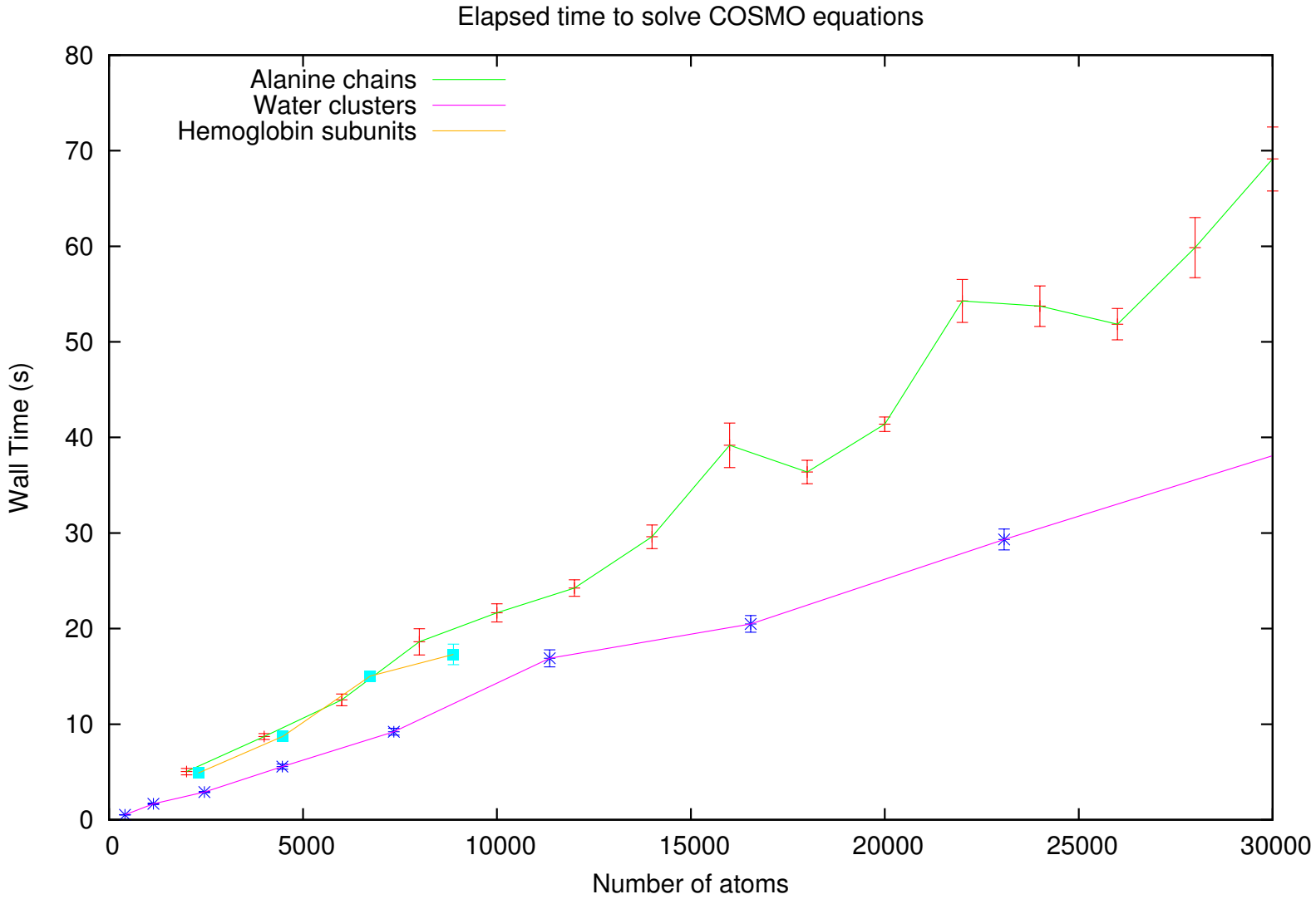
Summary

- A truncated spherical harmonics basis set can be used to discretize the problem
- Integrals are computed numerically in a very efficient way thanks to Lebedev-Laikov grids
- The $\mathcal{S}$ operator is diagonal in such a basis
- Only the blocks corresponding to overlapping spheres are nonzero: the matrix is block-sparse
- Jacobi iterations can be used (and DIIS extrapolation) : massive parallel implementation is possible
- Linear scaling with respect to the number of spheres

Model parameters:

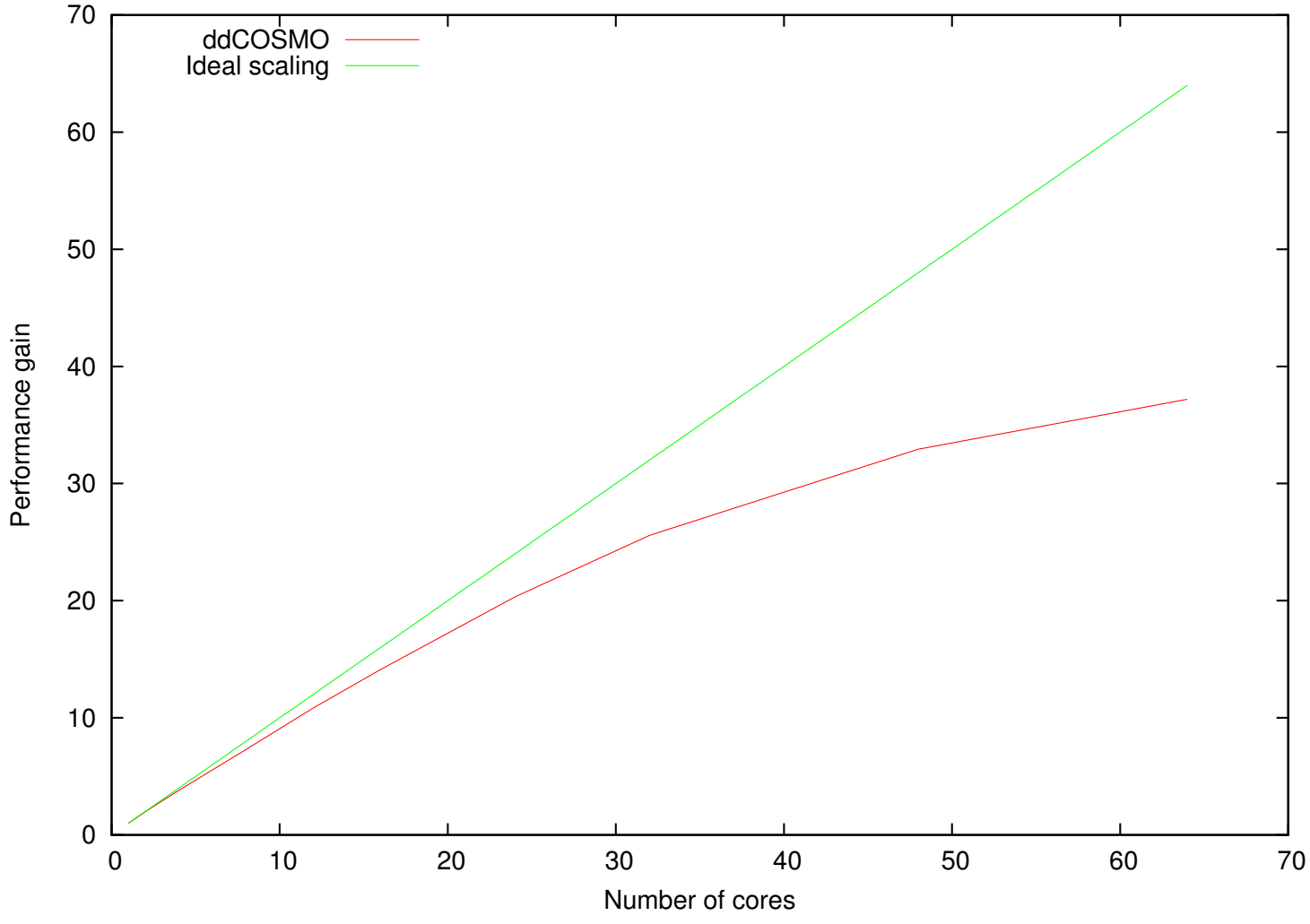
- Switching radius for the smoothing function
- Maximum angular momentum of the SH
- Number of integration points
- Convergence threshold

Globular and linear systems



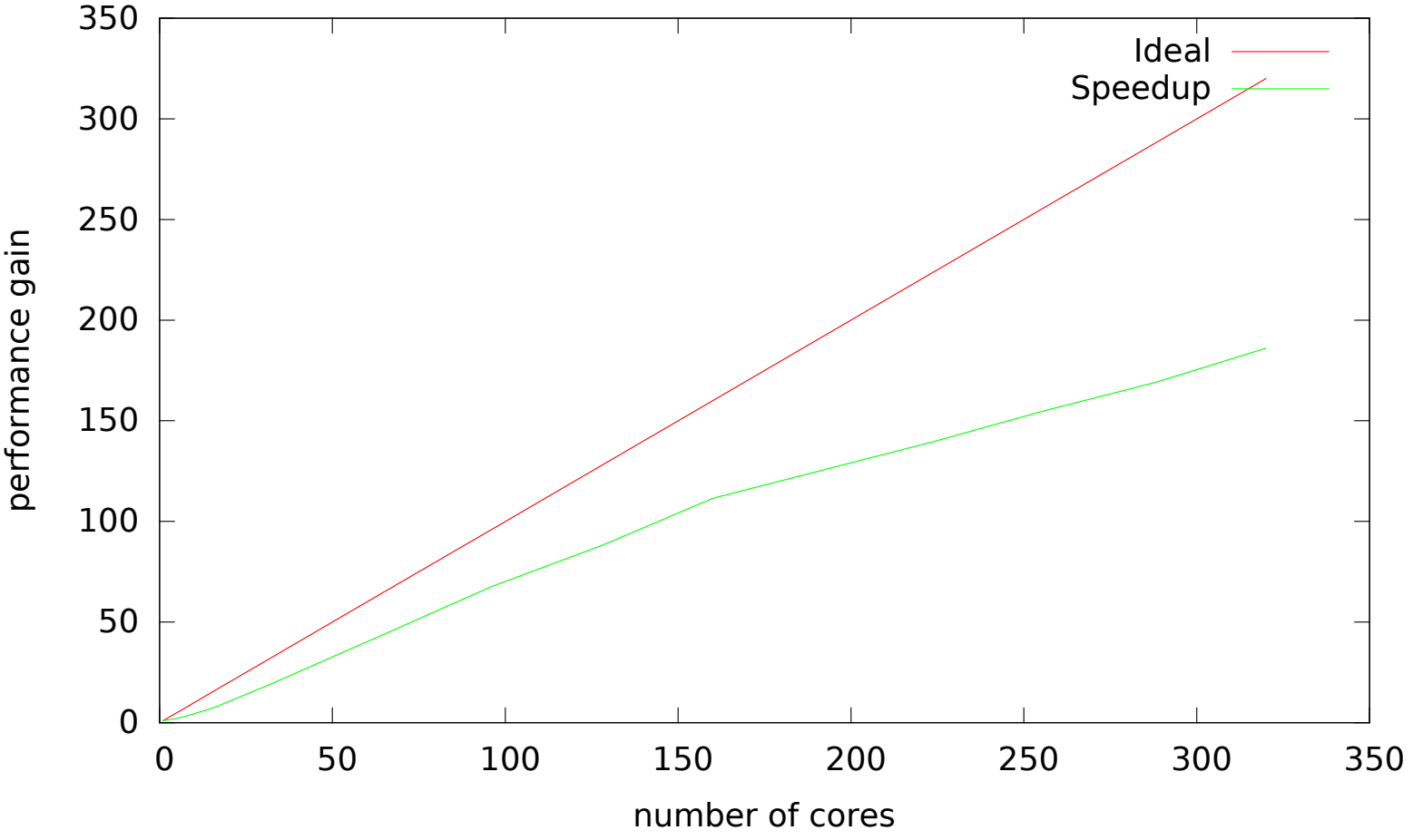
Shared Memory Parallel performances

- System of 10000 atoms



MPI Parallel performances

- System of 10000 atoms
- Covering communications with computations



Different discretizations and semiempirical methods

Table : Timings for the solution of the C-PCM/COSMO linear equations and for the computation of the Fock operator contribution for the different algorithms.

System	CSC - Iterative		FMM(old)		ddCOSMO	
	σ	F	σ	F	σ	F
Vancomycin	20”	1”	14”	1”	1”	1”
Hiv-1-GP41	1’26”	1”	22”	1”	1”	1”
l-Plectasin	1’35”	1”	29”	1”	1”	1”
Glutaredoxin	8’54”	1”	2’54”	1”	2”	3”
Glutaredoxin[1]	28’43”	-	-	-	-	-
UBCH5B	1h34’	4”	13’10”	3”	5”	8”
Carboxilase	12h57’	22”	1h27’	15”	9”	32”

[1] without the N^2 storage for CSC-Iterative

What has been done:

- Implementation in Tinker with classical force field (MD)
- Implementation in Gaussian (Semi-empirical, HF, DFT...)

But...

- In the case of polarizable FF, computation of the polarization energy becomes the bottleneck
- Goal : Study and improve the way this is done
- Then : Coupling with polarizable FF

Polarization Energy in polarizable molecular dynamics

- Static atomic multipoles : q_i, $\mu_{s,i}$, Θ_i
- Polarisability tensors (3*3) on atomic sites : α_i
- Induced dipoles on atomic sites : μ_i
- Global induced dipole vector (3N) : $\boldsymbol{\mu}$
- $T_{i,j}$ operators, $i \neq j$:

$$T_{ij}^{\alpha} = \frac{\partial}{\partial r_i^{\alpha}} \frac{1}{r_{ij}} = -\frac{r_{ij}^{\alpha}}{r_{ij}^3}, \tag{1}$$

$$T_{ij}^{\alpha\beta} = \frac{\partial}{\partial r_i^{\alpha}} T_{ij}^{\beta}, \tag{2}$$

$$T_{ij}^{\alpha\beta\gamma} = \frac{\partial}{\partial r_i^{\alpha}} T_{ij}^{\beta\gamma}, \tag{3}$$

In real life : damping on the T operators to avoid divergence : $\mathcal{T}$

Polarization Energy in polarizable molecular dynamics

$$\mathcal{E} = -\sum_{i=1}^{N} E_i^{\alpha}\mu_i^{\alpha} + \frac{1}{2}\sum_{i=1}^{M}[\alpha_i^{-1}]^{\alpha\beta}\mu_i^{\alpha}\mu_i^{\beta} + \frac{1}{2}\sum_{i=1}^{M}\sum_{j\neq i} T_{ij}^{\alpha\beta}\mu_i^{\alpha}\mu_j^{\beta}. \tag{4}$$

$$E_i^{\alpha} = \sum_{j\neq i} T_{ij}^{\alpha} q_j + T_{ij}^{\alpha\beta}\mu_{s,j}^{\beta} + T_{ij}^{\alpha\beta\gamma}\Theta_j^{\beta\gamma} \tag{5}$$

, the inducing field at each atom site

$$\mathcal{E}_{min} = \min_{\boldsymbol{\mu}\in\mathbb{R}^{3N}} \mathcal{E}(\boldsymbol{\mu}); \tag{6}$$

Polarization Energy in polarizable molecular dynamics

Corresponding linear system

$$\mathbf{T} = \begin{pmatrix} \alpha_1^{-1} & \mathcal{T}_{12} & \mathcal{T}_{13} & \dots & \mathcal{T}_{1M} \\ \mathcal{T}_{21} & \alpha_2^{-1} & \mathcal{T}_{23} & \dots & \mathcal{T}_{2M} \\ \mathcal{T}_{31} & \mathcal{T}_{32} & \ddots & & \\ \vdots & \vdots & & & \vdots \\ \mathcal{T}_{M1} & \mathcal{T}_{M2} & & \dots & \alpha_M^{-1} \end{pmatrix} \tag{7}$$

$$\mathbf{E} = \begin{pmatrix} \vec{E}_1 \\ \vec{E}_2 \\ \vdots \\ \vec{E}_M \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \vec{\mu}_1 \\ \vec{\mu}_2 \\ \vdots \\ \vec{\mu}_M \end{pmatrix}, \tag{8}$$

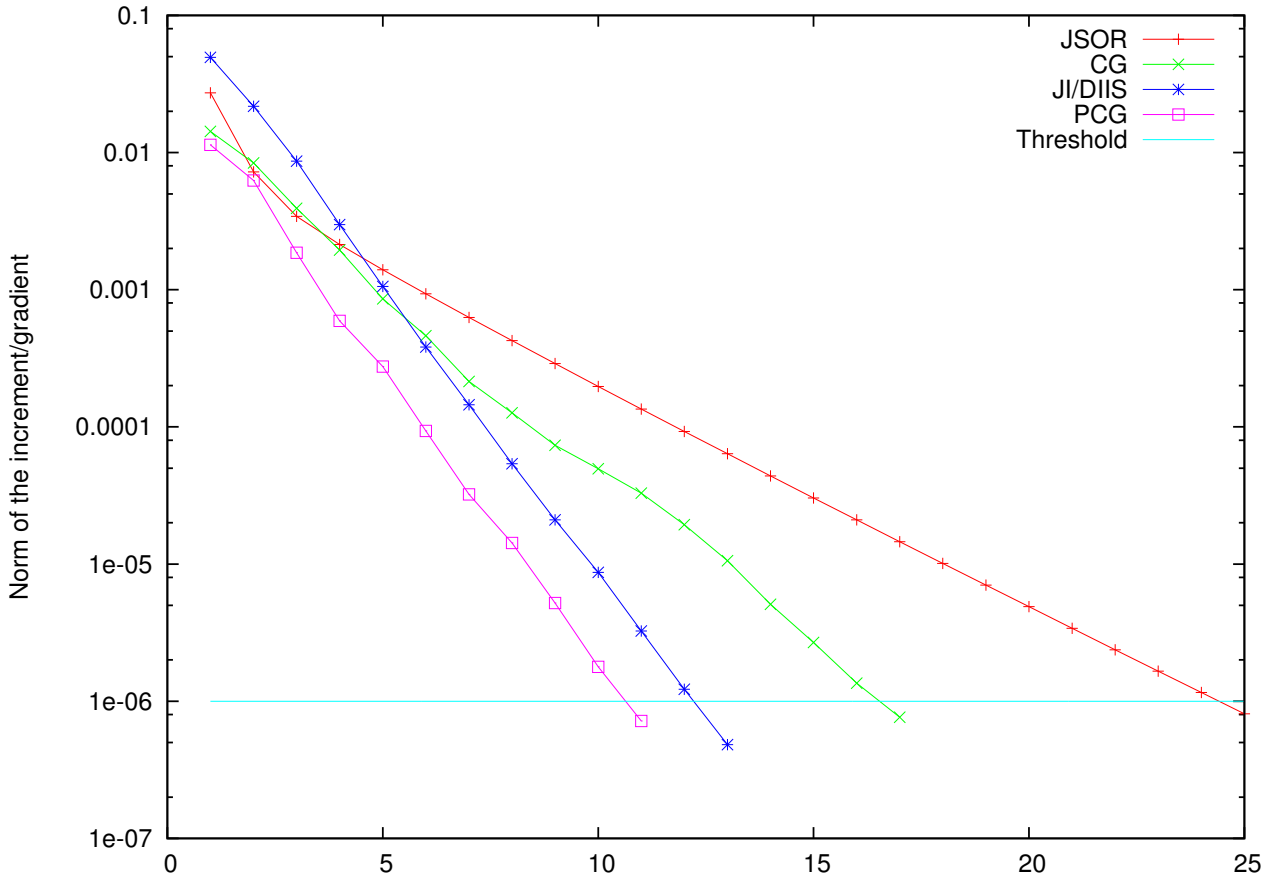
Polarization energy functional as

$$\mathcal{E}_d = \frac{1}{2}\boldsymbol{\mu}^\dagger \mathbf{T}\boldsymbol{\mu} - \mathbf{E}^\dagger \boldsymbol{\mu}. \tag{9}$$

$$\mathbf{T}\boldsymbol{\mu} = \mathbf{E} \tag{10}$$

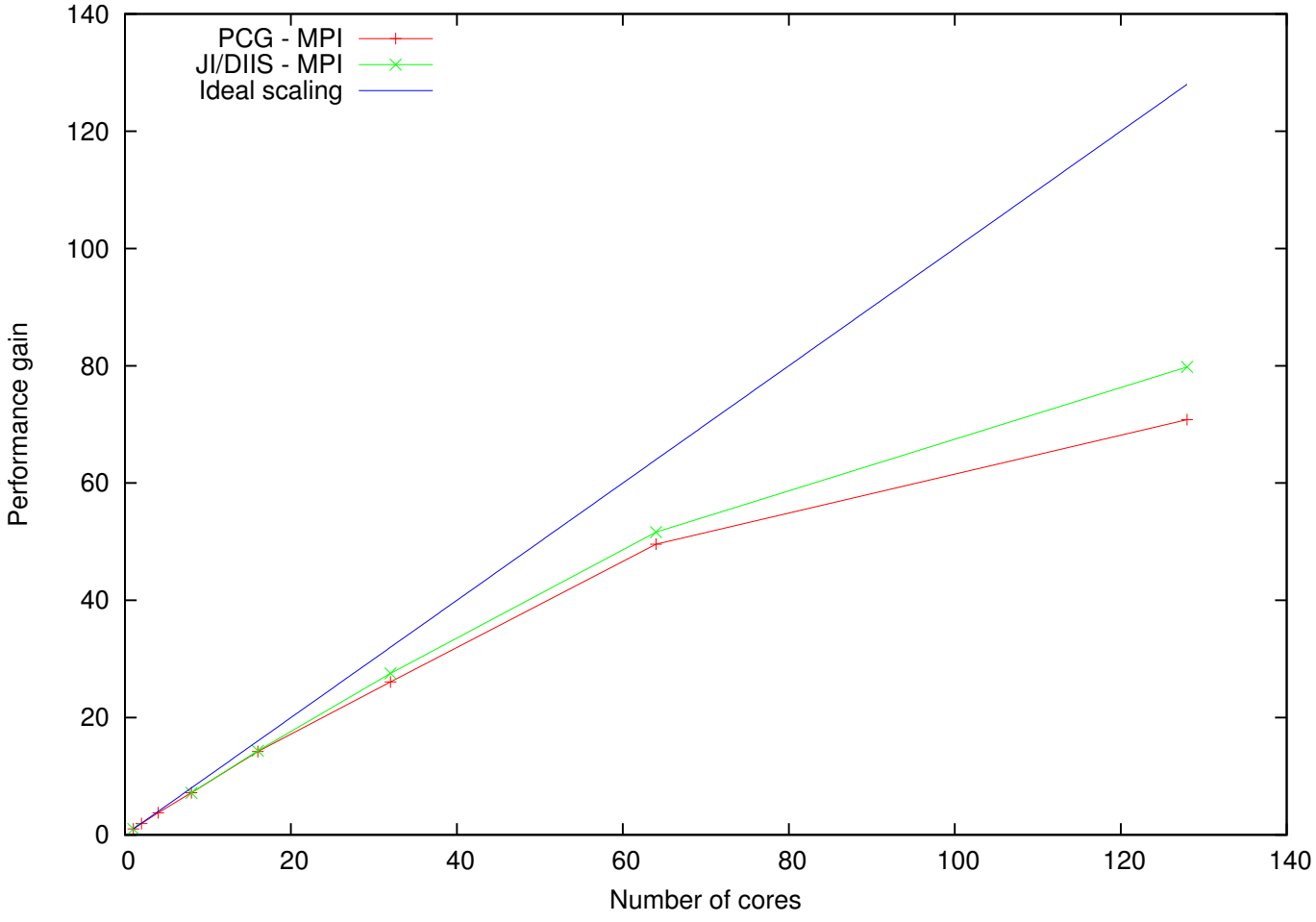
Numerical Results

- $\mathbf{T}$ is symetrical positive definite
- iterative methods : JSOR, Jacobi+DIIS, Conjugate gradient (with preconditionner)



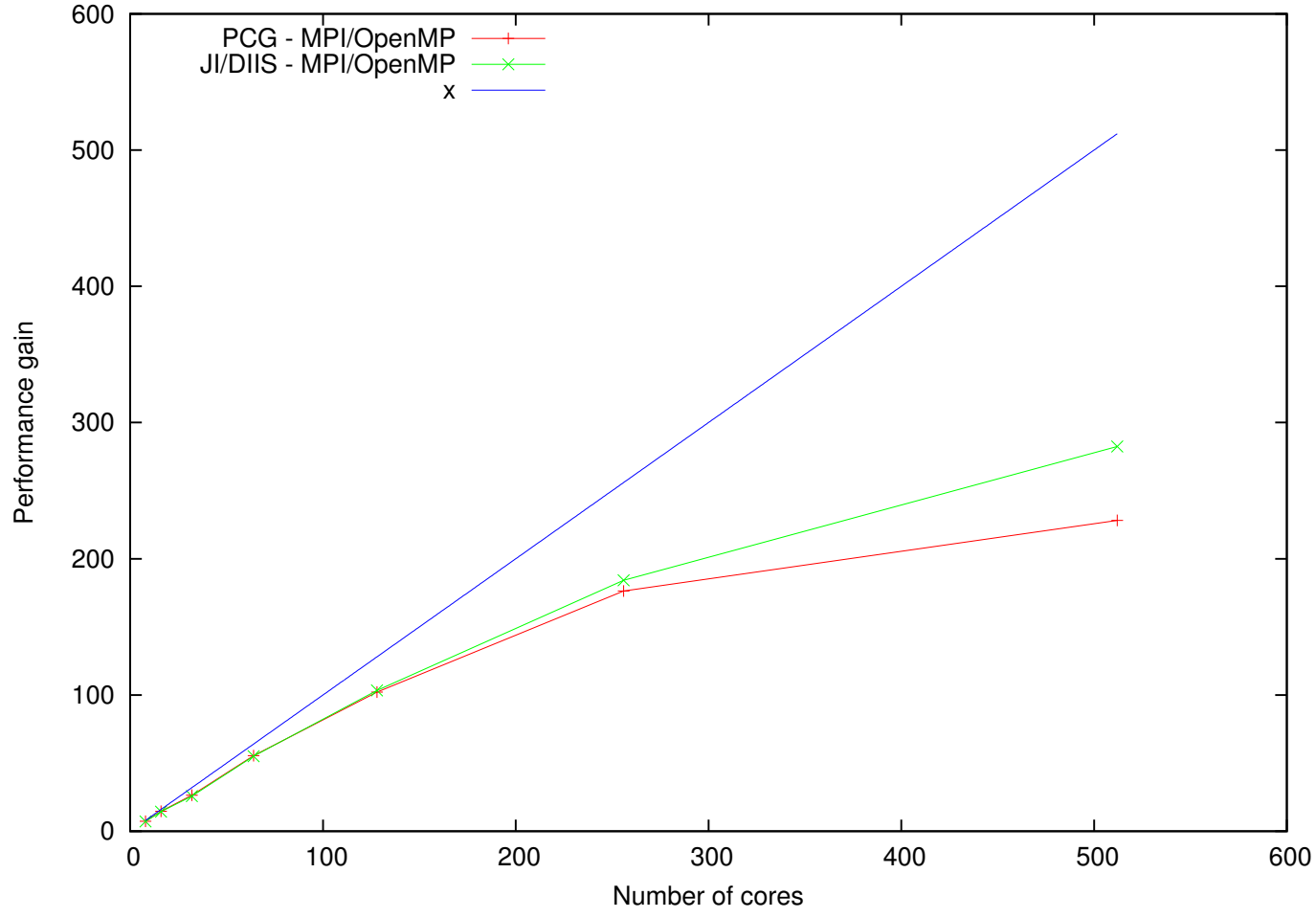
Numerical Results

- System of 10000 atoms



Numerical Results

- System of 10000 atoms



Coupling ddCOSMO/Polarization

Coupled Equations :

$$\boldsymbol{L}\sigma = g(0) + g(\boldsymbol{\mu}) \tag{11}$$

$$\boldsymbol{L}^* s = \psi(0) + \psi(\boldsymbol{\mu}) \tag{12}$$

$$\mathbf{T}\boldsymbol{\mu} = \mathbf{E}(0) + \mathbf{E}(\sigma, s) \tag{13}$$

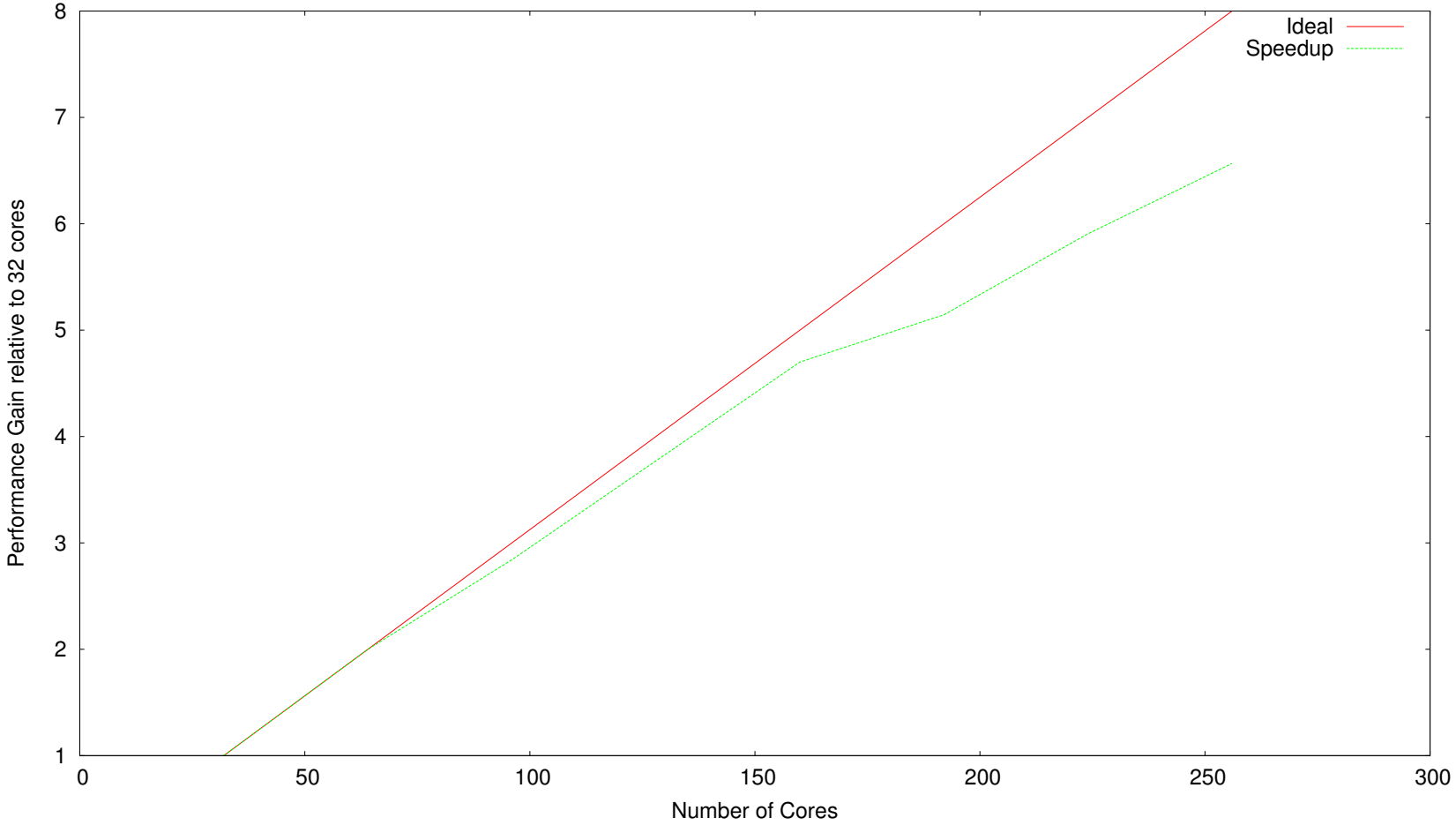
Gauss-Seidel like algorithm, at each iteration :

$\boldsymbol{\mu}_n, \sigma_n, s_n \rightarrow \boldsymbol{\mu}_{n+1}, \sigma_{n+1}, s_{n+1}$

- solve completely direct and adjoint ddCOSMO linear systems :
  - $\boldsymbol{L}\sigma_{n+1} = g(0) + g(\boldsymbol{\mu}_n)$
  - $\boldsymbol{L}^* s_{n+1} = \psi(0) + \psi(\boldsymbol{\mu}_n)$
- make one Jacobi step + DIIS extrapolation on the dipoles
  - $\mathbf{T}\boldsymbol{\mu}_{n+1} = \mathbf{E}(0) + \mathbf{E}(\sigma_{n+1}, s_{n+1})$

Numerical Results

- System of 20000 atoms
- Speedup relative to 32 cores



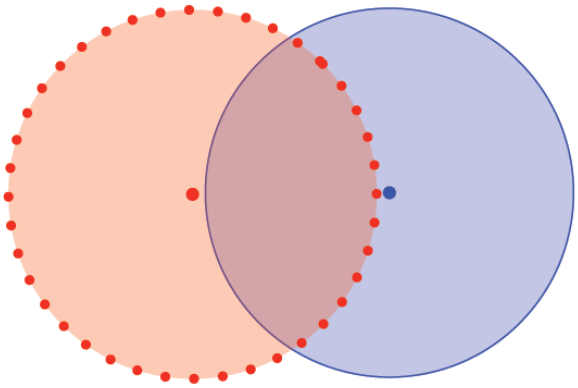
Conclusion and Perspectives

- Computational gain:
  - COSMO: 3 orders of magnitude
  - Polarization energy: at least 2 orders of magnitude (with parallelization)
- QM/MMPol/Solvent methods

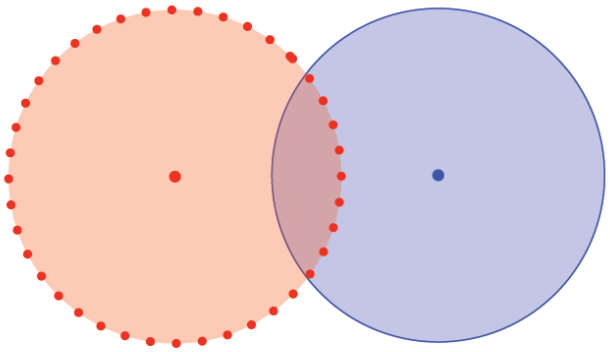
References

- *Fast domain decomposition algorithm for continuum solvation models: Energy and first derivatives*
  F Lipparini et al. Journal of Chemical Theory and Computation 9 (8), 2013
- *Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: I. Toward Massively Parallel Direct Space Computations*
  F Lipparini et al. Journal of Chemical Theory and Computation 10 (4), 2014
- *Quantum Calculations in Solution for Large to Very Large Molecules: A New Linear Scaling QM/Continuum Approach*
  F Lipparini et al. The Journal of Physical Chemistry Letters 5 (6), 2014

Regularization:

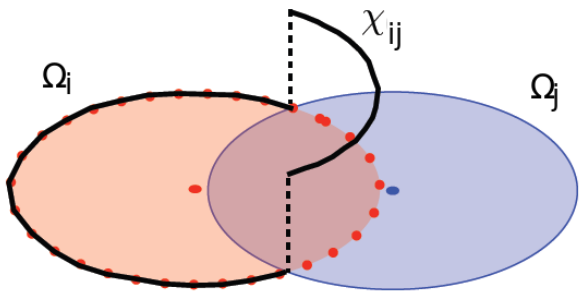


Regularization:



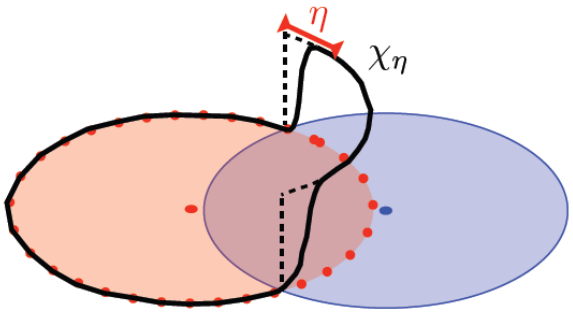
- Integration points buried into the cavity can suddenly become exposed!
- The boundary between internal and external is sharp (characteristic function!)
- Notice that without discretization, the energy would still be continuous.
- We can make the switching functions smooth

Regularization:



- Integration points buried into the cavity can suddenly become exposed!
- The boundary between internal and external is sharp (characteristic function!)
- Notice that without discretization, the energy would still be continuous.
- We can make the switching functions smooth

Regularization:



- Integration points buried into the cavity can suddenly become exposed!
- The boundary between internal and external is sharp (characteristic function!)
- Notice that without discretization, the energy would still be continuous.
- We can make the switching functions smooth